DOCUMENT RESUME

ED 322 174 TM 015 308

AUTHOR Campbell, Patricia B.
TITLE The Hidden Discriminator: Sex and Race Bias in Educational Research.
INSTITUTION Women's Educational Equity Act Program (ED), Washington, DC.
PUB DATE 89
NOTE 50p.; For related documents, see TM 015 304-307.
AVAILABLE FROM Women's Educational Equity Act Publishing Center, Education Development Center, 55 Chapel Street, Newton, MA 02160.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Data Interpretation; *Educational Research; Elementary Secondary Education; Ethnic Stereotypes; *Evaluation Utilization; Guidelines; Mathematics Achievement; Minority Groups; *Racial Bias; Racial Relations; Research Design; Research Methodology; Research Problems; *Research Utilization; *Sex Bias; Sex Differences; Sex Stereotypes

ABSTRACT

This monograph examines educational research and some of the myths surrounding it. The results of educational research strongly influence education. Much research, past and present, is biased. It is necessary to understand that race and sex bias can affect research, to understand how to determine whether bias is present, and to know how to minimize its effects. To demystify research, people need to gain familiarity with research terms and methods and its strengths and weaknesses, and they need to use their understanding to examine many of the myths that exist about research. The myth that research and researchers are always objective is the most difficult one to refute, but examples of bias are not hard to find. It is not difficult to find bias in historical educational research, but the bias of today tends to be more subtle. Two areas in which bias has strongly influenced research are student interracial interaction and sex differences in mathematics ability. Guidelines are provided for determining bias in research, reducing the negative effects of bias, and making research more valuable to policymakers and practitioners. Suggestions are also offered for improving research by viewing it more critically. An appendix, "A Beginner's Guide to Educational Research," introduces the basics. A 104-item list of references is included. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILABLE

TM015308

2

The Hidden Discriminator

Sex and Race Bias in Educational Research

The Hidden Discriminator

Sex and Race Bias in Educational Research

Patricia B. Campbell

Groton Ridge Heights
Groton, Massachusetts

Women's Educational Equity Act Program
U.S. Department of Education
Lauro F. Cavazos, Secretary

Discrimination Prohibited: No person in the United States shall, on the grounds of race, color or national origin, be excluded from participation in, be denied the benefits of, or be subjected to discrimination under any program or activity receiving Federal financial assistance, or be so treated on the basis of sex under most education programs or activities receiving Federal assistance.

The activity which is the subject of this report was produced under a grant from the U.S. Department of Education, under the auspices of the Women's Educational Equity Act. Opinions expressed herein do not necessarily reflect the position or policy of the Department, and no official endorsement should be inferred.

Printed and distributed by WEEA Publishing Center, 1989
Education Development Center, Inc., 55 Chapel Street
Newton, Massachusetts 02160

Contents

Acknowledgments vii

Introduction 1

Educational Research: Why Bother? 3

Demystifying Research: Exploding Some Myths 5

Stereotypes and Biases in Research 7
- Bias and What Is Studied
- Bias and Previous Research
- Bias and How Research Terms Are Defined
- Bias and Who Is Studied
- Bias in Research Tests and Measures
- Bias and What We Learn from Research

Historical Effects of Biased Research 19
- Intellect and Race
- Schooling and Sex
- Biased Research Conclusions: A Small Sample from the Past

Current Effects of Biased Research 23
- Student Interaction
- Mathematics Ability

Some Guidelines on Bias in Research 29
- Guidelines for Evaluating Research for Bias
- Guidelines for Reducing Bias When Conducting Research

Making Research Better: Next Steps 31

Appendix: A Beginner's Guide to Educational Research 33
- What Is Research?
- The Research Method
- Statistics

References 39

Acknowledgments

This monograph and the accompanying brochures owe much to the hard work and skills of the advisory board members:

Ms. Ann Leonard, independent consultant
Dr. Pamela Mason, Newton Public Schools
Ms. Georgess McHargue, author
Ms. Ellen Rubin, Educational Equity Concepts
Dr. Charol Shakeshaft, Hofstra University

Without their help, these materials would still be a dream.

The contributions of the field-test teachers, parents, administrators, counselors, and students should be acknowledged as well. Their comments and suggestions greatly improved the monograph and the brochures. Finally, thanks must be given to former Women's Educational Equity Act Program Project Officer Doris Shakin for her knowledge, support, and understanding.

Introduction

When we think of persistent barriers to educational equity, bias in research does not come readily to mind. In fact, it rarely comes to mind at all. Yet it should. The results of educational research strongly influence education. Much of what children read and how they are taught is based on research, and much research—past and current—is biased. Societal values, including attitudes about women and men, people of color and whites, can affect all members of society, including researchers; biased values can cause research results to be incomplete, exaggerated, and, in many cases, just plain wrong.

Research is used to develop and evaluate educational programs and materials, but if the research is biased, the information used in decision making is probably wrong. We need to be aware that race and sex bias can affect research, learn how to determine if bias is present, and, if it is, know how to minimize its effects.

"We" includes everyone involved in education—teachers, parents, administrators, counselors, and even students. Since all of us use, or are affected by, educational research results, we all need to know more about research and its strengths and weaknesses.

This monograph examines educational research and some of the myths that surround it. It covers many ways that bias can affect research, including its influence on who or what is studied, how the study is done, and what conclusions are drawn. Using examples from the past and the present, this monograph examines some effects of biased research and concludes with suggested guidelines for evaluating research and "next steps" to reduce the incidence and effects of bias on research.

Educational Research: Why Bother ?

We may wonder why we should become involved with educational research. The role that teachers, administrators, parents, students, politicians, and even the general public play in education is clear, but the importance of educational research in educational policy and practice is more difficult to see. Yet despite the apparent invisibility of educational research, its impact is pervasive.

Research results influence what is taught, how it is taught, and even the books and materials used (National Academy of Education, 1984). Research helps us assess everything from the relative effectiveness of different teaching methodologies to the relative value of computer-assisted instruction and computer simulations. For example, thanks to educational research, we have discovered the following:

- Class size is important. Students in very small classes do noticeably better than students in classes of "average" size, while students in large classes do slightly worse.
- Children learn math and science best when they use physical objects and do hands-on experiments.
- A good preschool experience has a strong positive influence on the long-term development of children at risk, particularly boys.
- Nonviolent methods of discipline, such as time out, are effective alternatives to corporal punishment.
- Computer-assisted instruction is an effective tool for basic skills remediation.

As the National Academy of Education concluded, "the findings of educational research generate principles and precepts for educational innovation and professional practice" (1984, p. 1). Good research can make education more effective.

Most of us have had little experience with research other than what we may have learned in an undergraduate or graduate course. This lack of experience contributes to a view of research as infallible and mysterious. Research is considered sacrosanct, rather than something to be evaluated and either accepted or rejected. Unless proven otherwise (and sometimes not even then), research is viewed as scientific, objective, valid, and good.

However, research, like everything else, can be bad or good, subjective or objective, accurate or wrong. Consider the following recent examples from reputable journals of research:

- Even though invalid, unreliable research is inaccurate research, Bordelon concluded, in a 1985 article published in *The Reading Teacher*, that "while some of the research does not follow good scientific principles for reliability and validity, much of it can help the classroom teacher." (p. 796)
- Without evidence, Collis, in a 1985 *International Journal of Women's Studies* article on sex differences in computer use, concluded that "females in this study appear to be their own enemies." (p. 213)
- Even though results were the same for disabled and able-bodied preschoolers, a 1985 *Journal of Educational Research* article by Fuchs et al. reported only the results for disabled preschoolers.

The factors that can cause research to be inaccurate or incomplete can be divided into two major categories: traditional sources of invalidity and societal biases.

Learning the traditional sources of invalidity and how to combat them is an important part of training researchers. Readers of research, too, need to be aware of these sources of invalidity, which include poor measurement techniques, a poor research plan, failure to account for differences among groups being studied (e.g., one group is older), and incorrect statistics. The Appendix, "A Beginner's Guide to Educational Research," provides an explanation of sources of invalidity as well as an overview of the basic components of research.

An examination of societal bias is not typically a part of the training that researchers or readers of research receive, even though the impact of bias can be devastating. Biased attitudes and stereotypic ideas about groups of people, based on their race, sex, or cultural background, greatly affect research.

Researchers are not immune to the influence of American racism (Thomas and Sillen, 1972); neither are they exempt from societal beliefs about the differences between women and men. Philosophers of science have discovered that scientific objectivity quickly becomes subjective under the influence of intense feelings (Nagel, 1961). As Russell concluded: "As soon as any strong passion intervenes to warp the expert's judgment, he [*sic*] becomes unreliable, whatever scientific equipment he may possess" (1959, p. 276).

Demystifying Research: Exploding Some Myths

Before using any research, readers must have confidence in the quality of the research and in their own ability to assess that quality. Research is neither infallible nor incomprehensible; you do not have to be a statistician to evaluate it. If a piece of research doesn't make sense to you, it may be because the research just doesn't make sense. At the same time, this doesn't mean that research can, or should, be evaluated from a position of ignorance. There are some basics, such as those covered in the Appendix, that people need to know in order to begin to understand and evaluate research. To demystify research, people need to gain familiarity with research terms, methods, strengths, and weaknesses, and they need to use their understanding to assess all research.

An important step in the demystification process is to examine some of the myths that surround research and researchers, including the following:

I. If it's been published, it must be true.

The publishing process weeds out some, but not all, bad research. In addition, studies finding differences between groups are more apt to be published than those finding no differences. "Let the buyer beware" also applies to research.

II. Research results remain true over time.

While some results remain true through the years, many do not. The changes of the past twenty to thirty years can mean that much of what we once "knew" no longer holds true. This is particularly true in education. As expectations and educational environments change, so do characteristics of students and teachers.

III. Sex and race differences exist in education.

There are few cognitive differences between girls and boys. Where differences exist, they are much smaller than the differences found within groups of boys or groups of girls. This is true of comparisons by race as well. In comparisons by race and by sex, "*within* group" differences are always greater than "*between* group" differences. Statements about the "average" girl and the "average" boy are powerful but misleading. In addition, such statements can become self-fulfilling, with the reader confusing the average with the individual.

IV. The beginning and end are the only important parts of a research study.

While the introduction and the conclusions may be the most interesting and readable sections, they do not provide enough information to assess the quality of the research. Only by reading the whole study—including its design and results—can one determine whether the conclusions are backed up by what actually happened.

V. Research and researchers are "objective," uninfluenced by societal values or their own view of the world.

While most researchers try to keep their interests and biases removed from their research, doing so is very difficult, if not impossible. As people from mystery writer Amanda Cross to philosopher Thomas Nagel have concluded, researchers are not immune to the influences of the world around them: "Kate marveled, not for the first time, at the ease with which academics deserted the cause of scholarly disinterest when their own most cherished opinions were at stake" (Cross, 1984, p. 67); "It is not easy. . . to prevent our likes, aversions, hopes and fears from coloring our conclusions" (Nagel, 1961, p. 488). In addition, researchers are often unaware of their biases, particularly when they are reinforced by society.

VI. I could never understand research.

With a little knowledge of statistics, interested people can use their problem-solving and critical-thinking skills to understand and assess most educational research.

Stereotypes and Biases in Research

Of all the myths cited in the previous chapter, the myth that research and researchers are always objective is perhaps the most difficult to refute. We, as a society, have invested science and research with an aura of truth. It is difficult to accept that research and researchers are influenced by the world around them, including its biases and stereotypes.

Given the strength and pervasiveness of societal attitudes, we all hold some biased ideas. Stereotyping—the assigning of traits and abilities to people based on their sex, race, or cultural background—is a fast but generally inaccurate way of categorizing people. Seeing a tall, lean, young Black man, people often think "basketball player"; seeing a girl climbing trees, they think "tomboy." In research as in life, stereotypes sometimes are accurate but more often are not.

As a society and as individuals, we hold a variety of preconceptions, many of which are based on people's race and sex. Even before a child's birth, our expectations of the fetus are based on its perceived sex rather than on individual differences. Many people, for example, still believe that if a fetus kicks a lot, it's a boy, whereas if it's quiet, it's a girl.

Expressions such as "she thinks like a man" and "don't worry your pretty little head" indicate society's view of girls and women and their intellectual abilities and interests. Complimenting someone by saying "she thinks like a man" implies that women and men have different thinking processes and that men's thinking processes are better. "Don't worry your pretty little head" implies that women's reasoning processes are different, less logical, and less serious than men's.

The situation is similar with respect to race. Expressions such as "Latin lover" and myths about Black sexual prowess reflect societal attitudes about people of color. They combine with myths about rhythm and a "happy-go-lucky" people to reinforce an image of Black people as less serious, less intellectual, and more "earthy" than whites.

While the statements above may be considered "just expressions," they are expressions that reflect societal racism and sexism. They are also expressions that set up (and reinforce) expectations and influence research. For example, the expressions found in the preceding paragraphs contribute to many researchers' expectations that sex and race differences will be found in studies of intellectual areas. As researcher Stephen Issac cautions:

> What the researcher expects to see, where he [*sic*] directs his attention, what he ignores or forgets, what he remembers or records, and even the way he interacts with subjects to alter their own expectations and motivational states, all can influence the results to fit his preconceptions. (1975, p. 58)

In other words, his (or her!) preconceptions can affect what is studied, how it is studied, and what is concluded about the results.

Bias and What Is Studied

Traditionally, people of color and women of all races have not been considered as important as white men, and less research has been done using them as subjects. For example, as recently as 1980 the editors of the *Handbook on Adolescent Psychology* announced that there was not enough research on adolescent women to warrant a chapter, even a short one (Gilligan, 1986).

When research has been done on groups other than white men, it has frequently focused on women and people of color as deviant or as victims. For example, researchers have asked:

- "Are women feminizing our schools?" (Sexton, 1969) *not* "Are men masculinizing our schools?"
- "Are Black people as intelligent as white people?" (Bettleheim and Janowitz, 1964) *not* "Are our intelligence tests biased?"
- "What are the negative effects of maternal employment on children's academic achievement?" (U.S. Department of Education, 1983) *not* "What are the positive effects of increased family income and maternal intellectual stimulation on children's academic achievement?"

Just as "the lake in which one decides to fish predetermines the kind of fish one will catch" (Stetson, 1982, p. 65), the questions one asks influence the answers one gets.

Societal attitudes and bias determine an issue's "cultural significance," which in turn affects the value given to different research topics. For example, researchers working on topics related to women and people of color report the need to "balance" their work with research on more highly valued areas (Campbell, 1980). Reflecting the experiences of many researchers, Steinem concluded:

> My own work on theories of gender-based power was academically suspect as single-factor analysis while my neighbor's work on one man's military acts during one decade was thoughtful, scholarly and basic. (1980, p. 98)

For many years, topics considered "women's issues," such as childbirth and informal support networks, were little valued and rarely examined. When these topics were covered, it was frequently from a perspective based on limited ideas of the roles women and men should play. For example:

- Studies of teachers who want to remain in the classroom (predominantly women) as opposed to those who want to move from teaching to administration (predominantly men) are seen as studies of deviant behavior rather than studies of different levels of aspiration. (Shakeshaft, 1979)
- There is much research on the problems of female-headed households and single-parent families, but little research on problems of two-parent families. (Committee on the Status of Women in Sociology, 1980)
- Studies of work, with rare exceptions, include only paid work. The work done by so many women, such as unpaid housework, child rearing, and agricultural work, is not included in studies of the labor force or in the gross national product. (Oakley, 1977)

Not surprisingly, bias has been found in studies of people of color as well. Omission is a major problem; usually, for example, people of color are not included in general studies. And when people of color are studied, the emphasis has been on them as victims or "problems." (Blacks are the group most often studied in this regard, followed by Latinos. Other racial groups, including Asian Americans and Native Americans, are rarely, if ever, studied.) In 1968, Billingsley found that studies of Black families displayed a selective focus on the negative aspects. In 1972, Thomas and Sillen concluded that "seen narrowly as a victim, the Black man appears in the learned journals as a patient, a parolee, a petitioner for aid, rarely as a rounded human being" (p. 47).

Little has changed. In 1984, Mathews, after surveying the literature on people of color and mathematics, reported that "the emphasis has been on minorities that have been unsuccessful in math—on why minorities don't enroll in math rather than why those who continue do" (p. 170). She found little focus on successful Black and Latino math students, and no comparisons of successful and unsuccessful math students of color.

The situation is even more bleak regarding research on both race and sex. Reports Scott: "One is almost overwhelmed with the... intellectual void that exists among social science scholars concerning the life experiences of Black women" (1982, p. 85). And even less data have been compiled about women and girls from other racial groups (Scott-Jones and Clark, 1986).

Since research topics are determined in large part by societal ideas of what's important and what's not (or what's right and what's not), systematic gaps in the educational knowledge base remain and grow.

Bias and Previous Research

New research builds on previous research and theory. However, much previous research used men and the male experience as norms against which all experience was assessed. Consider Kohlberg's widely accepted theory of moral development.

His six stages of moral development were empirically derived from a longitudinal study (one done over a period of years) of eighty-four boys in the United States. Even though the group studied was limited, the stages were said to be universal. In addition, boys and men were generally found to be at higher levels of moral development than women and girls were (Kohlberg and Kramer, 1969). This finding was seen as a problem not with the stages but with the women and girls. When Gilligan (1980) studied moral development in terms of women's lives rather than trying to evaluate those lives in terms of male-based models, she developed a theory very different from Kohlberg's.

Models based on half (or less than half) of the human race are, by definition, incomplete. Achievement motivation, like moral development, is an area in which a theory developed and tested on men and boys was used to explain the behaviors of females as well (Atkinson, 1958). Almost twenty years passed before there was an effort to look at the theory in light of the realities of women's and girls' achievement motivation (McClelland, 1975). Yet despite the growing awareness that any theory based on males alone does not reflect the total human experience, there is still a general assumption that theories based on whites reflect the total human experience. When these theories are used as the basis for further research, their inherent bias is perpetuated.

If belief in a theory is strong enough, it takes a lot to shake it. For example, although for many years Cyril Burt was the "guru" of work on the genetic basis of intelligence, by 1958 many were aware that Burt had falsified his data in order to support his theory that heredity determines intelligence. Nevertheless, Burt's work continued to be cited in almost every psychology textbook published over the next twenty years (Hearnshaw, 1979; Lewontin, Rose, and Kamin, 1984).

Strong opinions about women and men, about people of color and whites, can result in theories flexible enough to support those biases. The topic of sex differences in brain hemispheres, in "brain lateralization," is but one example. One researcher concluded that where men tend to show greater lateralization, such as in spatial skills, greater lateralization seems to be correlated with greater ability. However, she *also* concluded that when women show greater lateralization, then that same greater lateralization may be correlated with less ability (Witelson, 1978).

Sometimes it appears that the old saying "don't confuse me with facts" should be changed to "don't worry, with my theory I can explain away facts."

Bias and How Research Terms Are Defined

One major way that bias influences research is in how terms are defined. Nowhere is this more true than in definitions of race. There are two major definitions of race—social and biological. A social definition of race is based on societal perceptions; in other words, if the society views you as Black, then you are. A biological definition of race is based on genetics, on the presence or absence of biological

characteristics unique to a specific race. These two very different concepts are often confused, and the same label is used for people with very different backgrounds. For example, children of one white and one Black parent are frequently classified as Black and studied as such. Historically, and to a great degree even today, one is defined as either "all white" or "not white." There is no standard, consistent way to define people by race, making it very difficult to determine what racial differences actually exist. Studies concluding a genetic or biological basis for Black/white differences in intelligence and achievement (e.g., Jenson, 1969) have used an individual's self-definition of race to determine racial classification. However, when social definitions of race are used, *no* conclusions about genetic or biological differences can be made.

Ignoring subjects' socioeconomic status can also affect findings on race. Most early studies of racial differences did not control for the effects of socioeconomic status, even though the white sample, reflecting society in general, generally had a higher socioeconomic status than the Black sample did (Pettigrew, 1964). Differences in behavior were, however, concluded to be racial; the influence of socioeconomic status was not even considered (Graves and Graves, 1978; Pettigrew, 1964). The situation has not changed much. Even in 1986, Scott-Jones and Clark reported that researchers were still not accounting for the impact of socioeconomic status on studies dealing with racial background.

Definitions of socioeconomic status can also lead to biased research. Until the 1970s and even occasionally today, the socioeconomic status of a woman was defined by that of her father (if she was single) or her husband (if she was married). Thus the brain surgeon married to the bricklayer and the waitress married to the truck driver were considered, for research purposes, to be from the same socioeconomic level, whereas the typist married to the accountant was at a higher level. This arrangement made cross-group comparisons of women using socioeconomic status very suspect (Nichols, 1978).

Bias affects research in other ways as well. One crucial factor is that the research done on people of color and on women is frequently not used in the design of other studies. Consider, for example, the following findings:

- Black students of low socioeconomic status were found to test better with Black testers, while for other Black students the race of the tester made little or no difference (Samuel, 1977; Sattler, 1970). Therefore, studies using white testers may exaggerate Black class differences.
- Boys have been found to exhibit more antisocial behaviors when an adult is present; girls' behavior does not change (Caplan, 1975). Studies of behavioral sex differences that ignore this information can lead to inaccurate conclusions and a reinforcing of stereotypes about boys' behavior.
- Boys perform better when someone is watching; girls perform better in cooperative situations than in competitive ones; and both sexes tend to act

more stereotypically in the presence of adults (Greenberg, 1978; Maccoby and Jacklin, 1974). Studies with adult observers or studies using competitive testing that do not account for these findings will be inaccurate.

Not controlling for such findings, particularly in studies of race or sex differences, means that what is being studied is more apt to be some combination of the testing procedure and sex and race than just sex or just race.

Yet even today, graduate students report that their professors tell them always to do separate analyses by sex, without telling them of the pitfalls and complexities of such analyses (Lipson, 1986). Analysis by sex without an awareness of the previous research on sex differences and without accounting for other possible areas of difference generally leads to inaccuracies.

Bias and Who Is Studied

Traditionally, men have been the population studied in research related to education and other social science areas. An analysis of samples in research published in education and social science journals found that 22 percent of the articles did not even give the sex of those studied; almost half of the articles that did note their subjects' sex involved only one sex, most frequently males (Campbell, 1981).

While the number of single-sex studies has been reduced, they are still being done and the sex most likely to be studied is still male (Ward and Grant, 1986). In addition, most longitudinal studies (those that continue for a number of years) began by collecting information on men and boys and have not been updated to include women and girls.

A somewhat different pattern emerges for people of color. While there have been a number of studies of "the Black experience," Blacks are only infrequently included in studies of human behavior (Stivers and Leckie, 1976). Other people of color are rarely studied at all. Most studies of human behavior do not even mention the racial breakdown of the subjects, even though further inquiries have found that such studies are generally based on white samples (Campbell, 1981).

Although most samples do not include girls and boys from a variety of racial and cultural groups, this fact has not stopped researchers from concluding that a study's results apply to them. Conclusions generalizing to all dyslexic children, when only boys were studied and no information about race was included (Frauenheim, 1978), or conclusions about the economic progress for Blacks when only Black men were studied (Smith and Welch, 1986), are common. A survey of educational research studies found that more than 90 percent of them overgeneralized their results (Campbell, 1981).

Since researchers usually generalize their results to "humans," and humans include both sexes from all racial and cultural groups, one might wonder why researchers would use single-sex or single-race samples to study the human

condition. Although many researchers don't acknowledge the bias in their samples, those who do give some interesting rationalizations for their choices. One researcher explained that he used only males "to avoid introducing an additional variable that might detract from individual and group examination" (Frauenheim, 1978, p. 22). Another researcher excluded Black subjects because they were rejected by their peers less than white students were and because rejection was what was being studied (Bryan, 1976). Asian Americans, Native Americans, and to a lesser degree Latinos are so infrequently included in research that most researchers don't even bother to justify their exclusion.

Researchers have given three major reasons for working with single-sex subjects: (a) "scientific," (b) practical, and (c) "extrascientific" (Prescott and Foster, 1974, p. 3). "Scientific" reasons given were that "sex differences were known to exist in the phenomena and the investigator did not wish to explore them" and "the theory being studied was restricted to one sex." The sex of those who were available to be studied and the need to keep the number of subjects "reasonable" were given as practical reasons, while the "extrascientific" reasons were that the use of one sex reduced the variability of the data and that the experiment "favored" the use of one sex. Research has not been done on why researchers have chosen all-white samples, but it is not unreasonable to assume that the explanations would be similar.

Including females as well as males and people of color as well as whites can make a study more complex but also more accurate. Single-sex studies *are* occasionally necessary, as in studies of childbirth, and are sometimes understandable, as in studies of football players (Ward and Grant, 1984). Single-race studies, however, are more difficult to justify. The rule should be that if the results are going to be applied to females and males from different racial/cultural backgrounds, then the study sample should include them. Research cannot be generalized to populations not represented in a study.

Bias in Research Tests and Measures

For a number of years there have been concern and debate about the impact of race and ethnic bias on achievement and aptitude testing. For example, as early as 1951, studies indicated that verbal IQ tests in English were not good measures of Spanish-dominant or bilingual children (Altus, 1953). This finding was, however, generally ignored.

That tests are written by white middle-class authors and standardized and normed on white middle-class students but used on poor children and children of color has long been recognized as a problem. It is, however, only recently that test developers and users have attempted to resolve it.

Additional issues exist. For example, a 1970 study concluded that social class was a more important factor than racial origin in predicting intelligence test scores

for white and Latino children in the United States (Christiansen and Livermore, 1970). Cultural background, geographic isolation, and low socioeconomic status often combine to provide the child of color with a frame of reference very different from test developers' (Anastasi and Cordova, 1953; Tyler and White, 1978).

Sex bias can also affect testing. Research suggests that girls do better on test items dealing with "stereotypically feminine" topics than on "stereotypically masculine" items covering the same skills. In the past, tests, particularly achievement tests, have generally included more "stereotypically masculine" than "stereotypically feminine" topics, thereby negatively influencing girls' scores (Coffman, 1961; Donlon, 1971). Based on such variables as the skill areas tested, the context in which test items are set (a birthday party or a football game), and the type of test item (a multiple-choice or an essay question), sex differences can be created or eliminated (Dwyer, 1976; Campbell and Scott, 1980).

Achievement tests still include proportionately more "stereotypically masculine" items, but many of today's standardized IQ tests have been carefully balanced to eliminate the sex differences found in earlier versions. Tests have been developed to support the assumption that females and males are equal in intelligence (Lewontin, Rose, and Kamin, 1984). There has been no similar assumption about whites and Blacks, Native Americans, or Latinos, and therefore tests have not been "balanced" for these groups.

Other examples of bias in testing include the following:

- Personality tests can overpredict psychological problems in people of color who, either by circumstance or by choice, are not assimilated into the mainstream culture. Incorrect answers are assumed to be based on ignorance of common values when they may be based on an opposition to those values or on differences in situations. (Cowan, Watkins, and Davis, 1975)
- A test that measures "need to achieve" gives pictures to subjects and asks them to make up stories about the pictures. Girls' lower number of achievement-oriented stories about females is seen as evidence of girls' lower need to achieve, rather than as an awareness that girls and women as a group are not generally permitted to achieve in our society. (Kaufman and Richardson, 1982)
- Self-reports of subjects' achievements and confidence do not take into consideration the possible effects of modesty and self-effacement; such traits are more likely to be taught to members of some groups (e.g., Asian Americans of both sexes and most women) than to others. (Kaufman and Richardson, 1982)
- Clothing can affect testing. For example, studies of children at play might in reality be measuring differences in playing in dresses and playing in pants. (Campbell, 1981)
- Observations can be affected by bias. Studies in which some observers were told they were observing a little boy and others were told the same child was a girl found the perceived sex of the child affected the observers' response. (Gerwitz and Dodge, 1975; Herman and Serbin, 1977)

The decision as to which standardized test is used in research can influence what is found. Thirty years ago it was found that differences in the intelligence test scores of bilingual students and Spanish-dominant students were much reduced when nonverbal IQ tests were substituted for verbal IQ tests (Anastasi and Cordova, 1953). Much more recently, we learned that sex differences in mathematics were significant if the Scholastic Aptitude Test: Math (SAT:M) was used but were minimal or nonexistent if the School and College Aptitude Test–Quantitative (SCAT–Q) was used (Benbow and Stanley, 1983).

Recently a major study concluded that Blacks have made great economic progress in the past forty years; the study was based on comparisons between the median income of the employed Black male and that of the employed white male. If measures such as per capita income, poverty rates, and employment rates were used—or even if Black women had been included in the study—the conclusions would have been quite different (Crawford, 1986; Smith and Welch, 1986).

Bias and What We Learn from Research

Many factors outside the research process can and do affect research results. One such example is the sex of the researcher. When researchers first examined 148 studies on how people are influenced, they concluded that women are more easily influenced than men. However, when they looked at the sex of the studies' authors, they found that male researchers were more apt than female researchers to find women more easily influenced. Similar results were found in studies of people's skills in understanding nonverbal behavior; female researchers were more apt than male researchers to find women better at decoding nonverbal behavior (Eagly and Carli, 1981).

In these two examples, researchers portrayed their own gender more favorably. One wonders what the results would be if there were equal numbers of female and male researchers—equal numbers of sex differences favoring women and men, or perhaps no sex differences at all? Similarly, one wonders what the results would be if the number of researchers of color were equal to the number of white researchers.

The sex of the researcher is not the only important "outside" variable. When the research was done may also be significant. In the past twenty years, the roles of women and men, people of color and whites, have changed tremendously; so have the attitudes and tools of many researchers. For example, the studies of girls' and boys' vocational interests conducted more than ten years ago used tests that treated girls' and boys' interests and aspirations differently. One of the best-known vocational-interest inventories, the Kudor, provided pink answer sheets for girls and blue ones for boys. Career suggestions were determined by sex; careers such as physician, airline pilot, and veterinarian were suggested for boys, and stewardess, hotel housekeeper, and nurse, for girls. Studies using these tests hold little relevance for today's students or counselors.

There are other, more complex examples of the effects of a study's age. A 1982 analysis of studies of sex differences in cognitive abilities found males generally outperforming females in quantitative, spatial, and articulation areas and females outperforming males in verbal areas. However, a strong correlation was found between the sex differences reported and the dates of the studies. The older the study, the more likely there were to be sex differences and the larger those differences were (Rosenthal and Rubin, 1982).

A 1978 review of individuals' skills in "reading" people found a similar correlation. The more recent the study, the more women's scores improved compared with men's (Hall, 1978). Thus the date of the study should be considered in evaluating it.

Yet another factor influences the accuracy of the research we read. The publication and dissemination processes favor the finding of differences, not similarities. Studies finding statistically significant differences are more likely to be submitted for publication and more likely to be published. Such studies are, in fact, even more apt to be completed than those with no significant differences. Researchers are more than four times as likely to give up on a problem if preliminary work reveals no significant differences (Greenwald, 1975; Smith, 1980). Thus a study finding sex or race differences is more apt to be published (and read) than a study finding no differences.

This "publication bias" means that there is a focus on difference, on one group being found higher, better, or more skilled than others. Since this is what we are most apt to read or hear, it is also what we tend to believe.

The factors cited above are important, but the most influential factor is the attitudes of the researchers themselves. As discussed earlier, what researchers believe can determine what they see and how they interpret data. An early example can be found in Yerkes' work on chimpanzee behavior, work often quoted as justification for "natural" sex roles. Yerkes concluded that male chimps were naturally dominant, female chimps naturally subordinate. However, Herschberger, using the same data, drew some very different conclusions. Speaking from the perspective of a female chimp, she wrote:

> When Jack takes over the food chute, the report calls it his "natural dominance." . . . While I'm up there lording it over the food chute, the investigator writes down, "the male temporarily defers to her and allows her to act as if dominant over him." Can't I get any satisfaction out of my life that isn't allowed me by some male chimp, damn it. (1948, p. 10)

Drawing conclusions based on attitudes of what is appropriate is not limited to studies of chimps. In 1885 a *Psychological Review* article concluded that whites' slower reaction time (compared with Blacks' and Native Americans') was proof that whites were the superior group (Gossett, 1963). In 1887 Romanes found that

women's better reading skills indicated a lack of "deeper qualities of the mind" (Tolbin, 1972, p. 49). There are more recent examples:

- A 1966 study of self-esteem found, to the authors' surprise, high Black self-esteem; they concluded, without evidence, that it was a defense mechanism against discrimination. (McDill, Meyers, and Rigsby, 1966)
- Studies of sex differences were found to be more likely to use the term *superiority* when the differences were in the men's favor than when they favored women. (Parlee, 1975)
- A "classic" study of infant behavior concluded that boys try to solve a problem, while girls give up. It did not mention the equally plausible conclusion that the boys and girls tried to solve the problem in different ways. (Goldberg and Lewis, 1969)
- An analysis of studies of Blacks found that most researchers (82 percent) "blamed the victim," concluding that when Black/white differences were found, negative differences experienced by Blacks were due to the individuals' shortcomings rather than suggesting other possible explanations such as racism. (Caplan and Nelson, 1973)
- A 1961 analysis of studies of girls' and boys' reading skills found authors more apt to conclude that a study was tainted or that a mistake had been made when a study did not find girls better readers than boys. (Coffman, 1961)

If researchers have strong expectations, they may include only the results that support those expectations or assume that results are "really" significant even when no significant differences are found.

The author of a recent of study of sex differences in mathematics concluded that girls are more apt than boys to make math mistakes, even though a table on the very same page as that statement showed no significant differences between girls and boys (Caplan, MacPherson, and Tobin, 1985). Similar results have been found in studies of sex differences in spatial skills. For example, males have been consistently reported as scoring higher than females in doing line mazes. Yet only 18 of 105 studies statistically compared female and male scores, and in only 4 of the 18 were male scores significantly higher than female scores (Caplan, MacPherson, and Tobin, 1985). This problem goes far beyond sex differences in mathematics. Other studies have contradicted their own results, concluding, for example, that people from father-absent homes feel more victimized and in less control (Pettigrew, 1964) and that Black students are more likely than whites to feel teachers do not like them (Brown, 1967).

It is important to realize that none of these researchers falsified their results; they were not trying to fool us. They reported the results that did not substantiate their conclusions, but their conclusions were greatly influenced by their own attitudes and expectations. Until societal expectations for people of color and

whites and for women and men are equal, until there is equal respect for both sexes of all races, then answers to questions about people of color and whites and about women and men will reflect our own prejudices.

Historical Effects of Biased Research

Since it is often easier to identify yesterday's biases than today's, let us begin by looking at the effects of biased research in the past.

Intellect and Race

Until very recently, researchers repeatedly "proved" the intellectual inferiority of men of color and all women. Measures such as brain weights, head sizes, and facial proportions were used to "prove" that Anglo-Saxons were highest on the evolutionary ladder, followed by Northern Europeans, Slavs, Jews, and Italians, with Blacks trailing far behind. (Other racial groups were generally ignored.) This ranking pertained only to males; Anglo-Saxon females were considered at the level of Black males, and few bothered even to categorize women from other backgrounds (Ehrenreich and English, 1979; Thomas and Sillen, 1972).

Morton, one of the early researchers in this area, measured the capacity of a small number of human skulls and concluded that Blacks and Native Americans had smaller brain cavities than whites did and thus were less intelligent. Like so many other researchers, Morton was so convinced of his hypotheses that he simply discarded the data that did not support his hypotheses: in his final calculations, smaller skulls belonging to people of color were included, while smaller skulls belonging to whites were "thrown out." Had the smaller white skulls been included, Morton would have found no differences in skull size (Gould, 1981).

Researchers in this area were very good at finding flaws in research conclusions that did not support their point of view. However, they routinely ignored the very same errors in research results with which they did agree (Gould, 1981).

From skull size, researchers moved to brain weight as a measure of intelligence. Again, expectations determined results. When no differences could be found between brain weights of Blacks and whites, some very creative reasoning was used to support the predetermined conclusion that whites were more intelligent. For example, most studies of brain weights used the brains of unclaimed bodies. When the results showed no differences linked to race, one researcher ingeniously concluded, with no evidence at all, that only the lowest class of whites (prostitutes and "the depraved") would become unclaimed bodies, while Blacks at all socioeconomic levels would be abandoned at death! Thus the findings of no difference were said to be based on the "fact" that the lowest whites were being compared with all Blacks and that the data "do perhaps show that the low-class Caucasian has a larger brain than a better-class Negro" (Bean, 1906, p. 409; Gould, 1981, p. 79).

The brain weights and thus, supposedly, the intelligence of white women and

men were also compared. Finding women's brains smaller than men's, researchers completely ignored the reality that women are generally smaller than men and concluded that women and men could not be treated equally until their brain weights were the same (Tolbin, 1972).

By the twentieth century, intelligence tests began to replace brain measurements. However, with few exceptions, the conclusions remained the same. Ignoring test developers' assertions that test score comparisons should be made only among children from similar backgrounds, Terman concluded that, in comparison with whites, a low level of intelligence was "very, very common among Spanish-Indian and Mexican families of the Southwest and also among Negroes. Their dullness appears to be racial"(1916, pp. 91–92). He decided that children of Spanish-Indian, Mexican, and Black parents "are uneducable beyond the merest rudiments of training. No amount of school instruction will ever make them intelligent voters or capable citizens in the true sense of the word. Judged psychologically, they cannot be considered normal."

Terman was not a member of the Ku Klux Klan but rather a well-respected psychologist and educator who had a great influence on education. (Indeed, his work on the gifted is still being used today.) Like so many others, Terman let his beliefs influence his research. He did not account for the effects of any cultural biases in the tests he was using. Neither did he look at differences in the education that poor Mexicans and Blacks were getting in comparison with the education received by middle-class whites. He did, however, explain away embarrassing exceptions to his theories. In one study, Terman found the intelligence scores of hobos "distressingly high." Since it would not do for hobos to have higher scores than "more respectable" people, Terman used only each group's lowest scores; the hobos, instead of being in the middle where they belonged, sank to the bottom.

In *The Mismeasure of Man*, a fascinating debunking of research on intelligence testing, Gould concluded:

> The history of scientific views on race serves as a mirror of social movement; . . . reflecting good times and bad; periods of the belief in equality and of rampant racism. . . . Changes in research findings on intellectual inferiority reflect changes in society, with biological determinism rising in times of political retrenchment. (Gould, 1981, p. 29)

Schooling and Sex

Research related to women and schooling has also served as a mirror of society. When society wanted women in the home, research "discovered" a scientific basis to justify women's remaining uneducated and at home. In *Sex in Education; or, A Fair Chance for Girls*, reprinted for seventeen editions between 1873 and 1972, Clarke concluded that higher education would cause a woman's uterus to atrophy. This amazing "finding" was based not on medical reports but rather on data that

educated women were less apt to have children than uneducated women were. The smaller proportion of married educated women and their greater access to birth control were not considered as an explanation for their lower fertility rates; these factors weren't even mentioned.

Female students were concluded to be pale, in delicate health, and "prey to monstrous deviations from menstrual regularity" (Ehrenreich and English, 1979, p. 128). When one study found proportionately more educated women than men in insane asylums, its authors concluded that higher education was driving women crazy (Bullough and Bullough, 1973). Even G. Stanley Hall (1905), one of the founders of modern psychology, wrote that the woman who used her brain "first lost her mammary function" and had little hope to be other than a "moral and medical freak" (Ehrenreich and English, 1979, p. 129).

These conclusions were made by researchers but were not based on research. No controlled research was done on the relationship between higher education and the physical loss of mammary function. Neither was research done on "monstrous deviations from menstrual regularity."

Today, when institutions of higher education are courting women students, research like this would not be condoned. The picture is different in athletics, however, where women's role is much less assured. Recent conclusions about how "serious athletic training" can negatively affect future maternity sound very similar to past conclusions about "serious education" and maternity. The more things change, the more they may remain the same.

Biased Research Conclusions: A Small Sample from the Past

Conclusions such as those which follow were used to support policies that set up dual educational systems based on race and denied higher education to women (Campbell and Klein, 1982).

> [The] scientific community has been blinded to the truth [of racial intellectual inferiority] by the duplicity of Franz Boas, Communists, Jews and Sentimentalists. (Garrett, 1961, p. 253)

> Blacks in spite of being bereft of a moral sense do have a great compensating gift. . . . [T]hey all sing. (Evarts, 1914, p. 340)

> All Negroes have a fear of darkness . . . are careless, credulous, childlike and easily amused. (Bevis, 1921, p. 69)

> [It is] well known that among the colored race there are many women who are supremely endowed with almost unique emotional equipment which makes their services ideal for infants and young children. (Gesell and Ilg, 1943, p. 273)

[The ovaries] are the most powerful agents in all the commotions of [women's] system; . . . on them rest her intellectual standing in society, her physical perfection. (Bliss, 1870, p. 96)

Current Effects of Biased Research

While it may not be difficult to see bias in historical research, it is often difficult to see it in today's studies. Yet societal biases are still alive and well and influencing educational research and policy. The following examples are but two of the educational areas in which societal biases have greatly influenced research.

Student Interaction

More than thirty years ago, "separate but equal" education was declared unconstitutional. Although education is still not fully integrated, the 1954 ruling did cause educational researchers to focus on integration and on interracial interaction among students (i.e., the degree to which students of color and white students would talk together, work together, and generally become friends). This work sought to find ways of reducing "racial isolation," and the results were used to develop multi-million-dollar programs to encourage or facilitate public school desegregation. The results are still being used to design current programs. Most of these studies were seriously flawed, however; their results were at best incomplete and at worst totally wrong.

We know that students are more apt to be with, talk to, and make friends with students of their own sex than to do so with members of the other sex. We also know that the interactions boys have with other boys differ from those girls have with other girls and differ as well from the interactions between girls and boys (Best, 1983). The few studies that have looked at both race and sex differences in student interaction suggest that boys are more apt to interact with boys from different races than girls are to interact with girls from different races (Schofield and Sanger, 1977). Yet most of the major studies of interracial interaction did not even indicate the sex of the students being studied, let alone look at any effects that the sex of the students may have had on results (Slavin and Madden, 1979; Weinberg, 1977).

Those studies that looked at race and sex tended to examine the interactions of girls with girls and boys with boys. The exceptions, studies that looked at the interaction of race and sex, found that sex, not race, was the best predictor of interaction. Upper-elementary students were more apt to talk with and work with same-sex students of a different race than with different-sex students of the same race (Campbell, 1980). Regardless of race, same-sex interactions were more positive than cross-sex interactions were. At the same time, girl-boy interactions were more apt to be negative than same-sex interactions were, regardless of race (Campbell, 1980). Sex, not race, was the important factor.

What has been reported as racial isolation may be racial isolation or it may be

that the students of color being studied were more apt to be of one sex and the white students of the other sex. Since girls have been found to have fewer acquaintances than boys and to be more likely to choose same-race girls as friends, differences in interactions may be related to the proportion of girls and boys being studied.

The interaction of race and sex is rarely studied or even considered, a fact that may be due to society's discomfort with interracial girl/boy relationships (Weinberg, 1977). Regardless of why the interaction of race and sex has not been studied, ignoring it has contributed to our lack of information about racial isolation in schools, the degree to which it exists, and what can be done to reduce it.

Many other variables that can affect student interaction have also been ignored. Social class, for example, has been found to be an important component of student interaction, at least at the high school level (Petroni, Hirsch, and Petroni, 1970). Thus if the socioeconomic status of Black students and white students is different—which in many integrated high schools is the case—then at least some of what is described as racial isolation may actually be attributed to class differences.

In addition, work on interracial interaction among students has been done primarily on Blacks and whites and generalized to others. It is not realistic to expect that relationships between Blacks and whites will be identical or even similar to those between (a) whites and other students of color, (b) recent immigrants and the native born, or (c) those whose native language is English and those whose native language is not. Yet when results of studies on Black-white student interaction are generalized to student interaction between whites and other students of color, that is exactly what is being done.

Because they were based on the available research, programs designed to encourage multiculturalism and cross-race interactions rarely considered the effects of sex and class in their design. Thus the programs are less effective than they could be.

Mathematics Ability

It is widely believed that boys are better than girls in mathematics, and many, including some educational researchers, believe there is a biological basis for any differences (Benbow and Stanley, 1980). Beliefs about gender and math ability play a large role in what research is done and what conclusions are drawn.

Reading research conclusions, or even the preceding paragraph, one would assume that the differences between girls' and boys' math abilities are large and extensive. This isn't the case. Some studies have found sex differences in various mathematical areas; other studies have not. And when differences have been found, they are usually small. Many girls have higher math skills than most boys do, and many boys have lower math skills than most girls do. Differences within groups of girls or groups of boys are much greater than differences between the "average" girl and the "average" boy. Yet this distinction is rarely made in the research or in

discussions of the implications of sex differences in mathematics for teaching strategies.

In the past, most studies of sex differences in mathematics either gave standardized math tests to large numbers of students or used student math scores that had been collected for other purposes, such as the SAT:M scores or those collected for the National Assessment of Educational Progress. Most of these analyses showed boys having higher scores.

A comparison of girls' and boys' test scores does not mean and cannot mean that being female (or male) "caused" the scores to be different. Since the researcher could not assign subjects to be girls or to be boys, no one can be sure that the only difference between the groups being studied was their sex. There are many differences in the experiences of girls and boys that could affect their math achievement, including differences in the number and types of math courses girls and boys take.

There are sex differences in math courses taken. Boys generally take more and higher level math courses than girls do (Becker and Jacobs, 1983). This factor, of course, affects the amount of math they know and their math test scores (Jones, 1986). Before taking a geometry course, boys in seventy-four schools were found to have better geometry skills than girls had. However, when students were retested after taking the course, no sex differences were found in geometry skills (Senk and Usiskin, 1983).

The effects of such variables as differential course-taking must be investigated before any conclusions can be drawn on the cause, or even the existence, of sex differences in mathematics. Yet, because so many researchers have been sure that sex differences exist and are "natural," differences in girls' and boys' experiences are rarely examined or even mentioned as a possible factor in sex differences. In fact, when the number of math courses a student has had is taken into account, sex differences in mathematics are reduced or eliminated (Jones, 1986; Pallas and Alexander, 1983).

In a very well publicized exception, Benbow and Stanley compared math achievement on the SAT:M for gifted female and male seventh-graders and found males scoring higher. Since female and male seventh-graders take the same math courses, the researchers concluded that the differences came from "superior male mathematical ability" and suggested a genetic/biological reason for the differences they found (Benbow and Stanley, 1980). Furthermore, although only academically gifted students were studied, Benbow and Stanley concluded, with no evidence, that their results would be observed even if a broader population were studied.

Benbow and Stanley's conclusions were reported in *Newsweek*, on NBC's "Today," and in newspapers nationally. From these reports many people, including educators and parents, concluded that sex differences in math achievement were genetic. On the one hand, mothers who had heard about the Benbow and Stanley study had lower expectations of their daughters' aptitude for and achievement in

mathematics; on the other hand, fathers who had heard about the study were *more* apt to think math was important for their daughters (Jacobs and Eccles, 1985).

While Benbow and Stanley did control for math courses taken, many other factors were not mentioned, much less taken into consideration. Even when girls and boys take the same math courses, their experiences and the encouragement they receive in class are quite different; boys, for instance, generally receive more praise and attention, which reinforces the concept of math as a male domain (Becker, 1981). Girls and boys have different math experiences and encouragement outside school as well. Boys' play experiences, for example, provide boys with more opportunities to develop and improve spatial skills (Greenberg, 1978).

The test used to analyze math ability also has an impact on research results. As indicated earlier, test developers have found that girls tend to score higher on essay and fill-in-the-blank questions, while boys tend to score higher on multiple-choice questions. Thus sex differences can be increased or decreased by changing the types of items on a test (Dwyer, 1976). Yet the type of test items is rarely considered, controlled for, or even mentioned in studies of sex differences in mathematics (or in other areas).

Test content also affects results. Many of the studies of sex differences in mathematics use the SAT:M. In 1971, Dr. Thomas Donlon of the Educational Testing Service stated that although males scored about 40 points higher than females on this test, the difference could be cut in half by having items cover subject matter that was more familiar to females. Today this sex difference could still be greatly reduced using the same technique.

Most researchers know that different math achievement tests find greater, lesser, or no sex differences. The Benbow and Stanley study used the SAT:M to measure sex differences even though the authors were aware that this test finds sex differences "both early and late," whereas other tests such as the School and College Aptitude Test–Quantitative do not detect an early sex difference (Benbow and Stanley, 1980, 1983). It is conceivable that they chose the SAT:M because they believed there were sex differences in mathematical ability. One wonders what an "objective" researcher would have done.

It is important, and possible, to account for differences in course-taking behavior; it is also possible to choose a test that minimizes sex differences in math. It is equally important but much more difficult to account for different treatment in the same math classes and for different experiences outside school. Whether or not such areas can be controlled in the research, readers need be made aware of them. Yet these issues and supporting research results have not been cited in national news magazines, nor have authors of studies in these areas appeared on national television. It may be that these results did not receive the publicity accorded Benbow and Stanley's work because the latter study reinforced the stereotype that boys are naturally better in mathematics, whereas the others challenged it.

Work on racial differences in mathematics is subject to similar caveats. The

relationship between the number of math courses taken and math achievement holds true for cross-race as well as for cross-sex studies. Studies have found the number of math courses taken to be a better predictor of math achievement than such variables as parent occupation, mean grade-point average, or racial/ethnic background (Jones, Burton, and Davenport, 1984; Welch, Andersson, and Harris, 1982). For example, studies of math differences between Blacks and whites frequently don't, but should, reflect the smaller number of Blacks taking advanced math classes and the fewer number of math courses they have taken.

There may be sex and race differences in math achievement, but at this point we don't know how large or how firm these differences are or what their causes are. Because of bias in research, it is difficult to know even whether the differences are real, let alone whether they are large enough to be of concern. In addition, without adequate information on causes, it is very difficult to determine the types of programs that need to be developed. Should efforts, for example, focus on affective issues (such as encouraging people of color and young women to take more math courses) or on cognitive issues (such as providing more experience in spatial skills)? Good research is a precursor to effective efforts to provide equal learning opportunities to all students.

Some Guidelines on Bias in Research

Yesterday and today, biased beliefs about female and male students and about students of color and white students have affected research results and thus influenced the educational decision-making process. The following guidelines may prove helpful in assessing research for bias, reducing its negative effects, and making educational research more valuable to policymakers and practitioners.

Guidelines for Evaluating Research for Bias

1. Can you tell the author's opinions or biases as you read the study? For example, the author's bias is clear when a study is done to determine the *negative* influence of mothers' employment on children's achievement.
2. Do authors use different words depending on the sex or race of those being studied? For example, if studies of father absence are labeled "father absence" while studies of mother absence are labeled "maternal deprivation," bias is present.
3. How is racial or ethnic group membership defined? For example, if genetic differences between Blacks and whites are concluded when no genetic definitions of Black or white are given, the study is biased.
4. Are the tests used "fair"? Does the study indicate whether the tests were developed and used with females and males from a variety of racial and cultural backgrounds?
5. Does the study describe who is being studied, including their sex and race?
6. Are the results of the study applied only to people like those studied or are they overgeneralized to include others? For instance, are people of color included in conclusions when only whites were studied?
7. Are sex and race similarities as well as differences reported?
8. Are the conclusions based on the author's results or on the author's expectations?

Guidelines for Reducing Bias When Conducting Research

1. The research design should account for confounding variables related to the race and sex of subjects. Researchers should determine the validity of measures of independent variables dealing with race and sex. Independent variables that define one group in terms of another, such as basing a woman's socioeconomic status on that of her husband or father, should not be used.

2. A review of the research literature should include a critical analysis of prior research, with information on *important characteristics* of the groups being studied. Researchers should expand their literature search to include publications focusing on women and people of color (e.g., *Psychology of Women Quarterly* and *Black Education*). Any literature cited in a research article should be subject to critical assessment that includes guidelines on research bias. Mention should be made of any major weaknesses.
3. Unless there is a demonstrable rationale for restricting a sample to one race or sex, samples should be multiracial and include both females and males. In the development and testing of models, researchers should include women and people of color, rather than using them, post hoc, to investigate how well they fit existing models devised from white male samples. If samples are not multiracial or do not include females and males, then a justification for the makeup of the sample should be made and the results should not be generalized to groups unrepresented in the sample. A sample's race and sex characteristics should be described in any report of the research.
4. Only tests that are not biased for or against the groups being studied should be used. In developing or selecting tests for research, researchers should avoid tests that
   - use exclusionary language or other offensive language or questions
   - do not include materials relevant to women and people of color
   - give no evidence of validity for the individual groups being tested
5. Researchers should control for possible effects of observers' perceptions of "appropriate" behaviors for subjects from different racial and gender groups. If possible, researchers should mask the sex of young subjects. Observers should be made aware of the effects that stereotypic expectations may have on their ratings. Sex and race differences found through observation should be substantiated using other methods of data collection.
6. Conclusions should be referenced directly to the results of the study. Nonstereotyped as well as more traditional explanations of results should be explored. If, for example, nonsignificant differences are found, they should not be reported simply as differences, and when sex and race differences are found, a variety of possible explanations for them should be considered.

Making Research Better: Next Steps

1. Apply the suggestions and guidelines included in this monograph to your own actions.
   - Don't make decisions based on what "research says" until you check the results for general accuracy and bias.
   - Don't pass on "facts" without checking on their accuracy.
   - Read the entire research study instead of just the beginning and end.
   - Use the same criteria to evaluate studies whose results you feel must be right as you do to evaluate studies whose results you feel can't be right.
2. Make others more aware of bias in research and its effects on education and other areas. Discuss issues related to bias in research with your students and colleagues. Consider distributing copies of the brochures that accompany this monograph (available from the WEEA Publishing Center, EDC, 55 Chapel St., Newton, MA 02160). These brochures were designed specifically for teachers, administrators, counselors, students, parents, and others interested in education.
3. When you find a study that is biased, do something about it. Write down what aspects of the study you think are biased and why; note what effect you think this bias may have had on the study's results. Write to the editor of the journal or magazine in which the research was published and describe your concerns. Send a copy of your letter to the author as well. Suggest that the journal include information on sex and race bias in research in the guidelines it provides to its authors and reviewers. Offer to become a reviewer yourself.
4. Professional organizations are paying more attention to standards and guidelines to increase the quality of research and evaluation. Some guidelines, such as *Standards for Evaluation of Educational Programs, Projects, and Materials* (The Joint Committee on Standards for Educational Evaluation, 1981), do not deal with issues of bias at all; other guidelines, such as those of the American Psychological Association, cover the effects only of sex bias on research; and still other guidelines, such as those of the American Educational Research Association, include the effects of both race bias and sex bias on research. Find out what steps, if any, your professional organizations are taking to address bias in research, and encourage them to develop, approve, and use guidelines to reduce bias in research.

5. Find out more about bias in research. The following references can provide additional information:
   - Campbell, P.B. (1983). Racism and sexism in research methods. *Encyclopedia of Educational Research*. New York: Macmillan.
   - Ehrenreich, B., & English, D. (1979). *For Her Own Good*. Garden City: Anchor.
   - Gould, S.J. (1981). *The Mismeasure of Man*. New York: W.W. Norton and Company.
   - Thomas, A., & Sillen, S. (1972). *Racism and Psychiatry*. New York: Brunner/Mazel.
6. Find out more about research in general. The Appendix, "A Beginner's Guide to Educational Research," is a good place to start.

Appendix

A Beginner's Guide to Educational Research

What Is Research?

Research, according to Webster's, is an "investigation or experimentation aimed at the discovery and interpretation of facts." In education there are two major types of research: *basic research*, in which the goal is a better understanding of learning and the educational process, and *applied research*, which focuses on finding information that will improve current educational practice.

The Research Method

Educational research, like research in other areas, relies heavily on what is known as the scientific method. This is a system of investigation that typically involves the following steps:

- development and statement of a question or a problem to study
- formulation of a hypothesis (a "best-guess" answer to the question being studied, based on existing theory and research)
- development and implementation of a structured plan (or design) to test the accuracy of the hypothesis or to answer the questions posed
- determination of the results of the plan
- generation of conclusions based on research results and the development of further research questions based on both the results and the conclusions

The Sample

In research, those being studied are called participants, or *subjects*. A group of subjects is called the *sample*. The sample is supposed to be representative of a *population*, a larger group to whom the results of the research can be applied (i.e., the results of a study on a sample are generalized to the population that sample represents). For example, if you selected ten students from each of ten classes for a research project, each of the students would be a subject, the one hundred students would be your sample, and the ten classes would be the population to which your results could be applied.

The Design

Both basic research and applied research can be done using one of several different designs, or plans of action. These plans are based on a variety of factors, including the type of research being done and the resources available.

The Experimental Design. If one wishes to determine whether something "causes" something else, the experimental design is the most effective. In this design, one or more groups, called the *experimental group(s),* receive some sort of treatment (e. g., a new reading program, a new tutoring program, or smaller class size) while a similar group, the *control group,* receives no treatment. The essence of the experimental design—that which makes it the "best" design for causal research—is that each of the subjects studied has an equal chance of being selected for either the experimental or the control group. Use of this design increases the chance that the only difference between the experimental group and the control group will be that one receives the treatment and one doesn't. Thus if differences show up between the experimental and control groups, those differences can be said to be caused by the treatment. For example, if the one hundred students in our sample are selected by chance to go either into a group that receives money for getting an A or into a group that receives no money for getting an A, then we have an experimental design. If the subjects receiving money get higher grades than those who don't receive money, then, because it is an experimental study, we can say that, for that sample, receiving money for good grades improves student grades.

The Ex Post Facto Design. Experimental designs are not always appropriate. For instance, because people cannot be randomly assigned to be female or male or to be Black or white, an experimental design cannot be used to study sex or race. Neither can an experimental design be used to study something that has already occurred, because it is then too late to randomly assign subjects.

For these kinds of studies an *ex post facto,* or *quasi-experimental* design can be used. *Ex post facto* is a Latin expression meaning "after the fact." In an ex post facto study, the researcher does not control who is in the experimental group and who is in the control group. Therefore, it is not possible to be sure that the only difference between the groups being studied is the treatment or to conclude that the treatment "caused" any differences in the group.

For example, a researcher studying young children at play might find that girls and boys have different patterns of play. The researcher could conclude that girls and boys play differently but could not conclude that being a girl or being a boy "caused" the children to play differently. There are many other variables that might account for the differences. These variables might include that all the boys are wearing pants while a number of the girls are wearing dresses; that teachers frequently give different instructions and play suggestions to girls versus boys; or

that parents are more apt to be concerned about girls "keeping clean" than about boys doing so.

Post Hoc Fallacy. Drawing invalid conclusions based on ex post facto research is so prevalent that researchers have a special name for it: a *post hoc fallacy*. Since by definition all research comparing females with males, people of color with whites, and disabled with able-bodied persons is ex post facto, it is particularly important to check for post hoc fallacies in such studies. You should suspect the existence of a post hoc fallacy whenever a study concludes that being female or Black or disabled, for example, causes something to happen, whatever that something might be.

Other Designs. There are a number of other ways that research can be done, including the following:

- *Survey, or descriptive, research,* in which there is no treatment and subjects respond to a series of written or oral questions describing a situation or area of interest. A study of student attitudes toward school would be an example of survey research.
- *Qualitative, or naturalistic, research,* in which the researcher observes people in a natural setting and, over a period of time, almost becomes a part of a group in order to be able to analyze group processes and interactions. A study of how fourth-graders' behavior changes in terms of how fourth-graders interact with the teacher during the school year would be an example of a qualitative study.
- *Correlational research,* in which the degree to which changes in one variable are reflected in changes in one or more other variables. A study of the relationship between achievement test scores and grades would be an example of correlational research.
- *Historical research,* in which analysis is based on documents and data from the past. A study of the different ways that reading was taught in the nineteenth century would be an example of historical research.

It is important to note that regardless of the research design used, *if there is no random assignment of subjects to the treatment, then you cannot be sure that the treatment caused any differences.*

Sources of Invalidity

Obviously, the quality of research—its validity, the degree to which results are accurate and can be attributed to that which is being studied—is very important. Researchers have long been concerned about the validity of their work and have

attempted to design studies that control for as many sources of invalidity as possible, even though few control for or even consider societal biases. The following is a list of the more common sources of invalidity not related to societal biases.

- *The Hawthorne Effect.* Being studied and getting extra attention, being "special," may be enough to cause changes in the subjects independent of what is being studied. The original Hawthorne study was done with factory workers. Researchers found that productivity increased when they did positive things (increased light, increased breaks); they also found that productivity increased when they did negative things (increased room temperature). Further work found that it was the increased attention that raised productivity. Using a second group that gets the attention but not the treatment controls for the Hawthorne Effect.
- *Maturation.* Just growing older can have a strong influence on subjects, particularly if young children are being studied. For example, if researchers are studying the effects of a year-long program on children's language development, they must remember that children's language skills will improve in a year regardless of the program used. Without a same-age control group, the researcher will not know how much of a change is caused by that which is being studied and how much is caused by the subjects' getting older.
- *Testing.* Testing can affect a study in many ways. Obviously, if a test doesn't measure what it is supposed to measure, then results will be incorrect. In addition, taking a test can affect subjects; changes in subjects may be due to the test rather than the treatment. For example, the practice of taking a pretest on fractions might do more to increase students' abilities to work with fractions than the treatment does. Finding tests that have been found to be valid (that do measure what they say they measure) and using a control group that takes the tests but not the treatment are ways of controlling for the influence of testing.
- *History.* In an ideal study, the only difference between groups being studied is the treatment; however, during a study, groups may have different experiences (a teacher might get sick, a school might start a new project). By the end of the research period, the different histories of the groups, rather than that which is being studied, might be the cause of any changes. It is very difficult to control for history; being aware of the unexpected and unintended events that occurred and reporting them in the results are about all that can be done.

Statistics

Wait! Even though your first impulse may be to skip this section, *read on.* Many of us are afraid of statistics and convinced that we can never understand them. That

does not have to be the case. Even though most of us will never become theoretical or even applied statisticians, we can, with a little effort, learn enough to begin to make sense of the statistical section of a piece of research. Always keep in mind that statistics are just a way of reducing large amounts of information (test scores, height, attitudes, rankings, almost anything) into summaries that provide useful information.

There are two basic types of statistics: *descriptive* and *inferential*.

Descriptive Statistics

Descriptive statistics as their name implies, reduce and describe a large amount of information. Typical descriptive statistics include the *mean* (average), *mode* (most frequent score), and *median* (score at which half the scores are below and half are above). The *standard deviation* is the measure of how varied or spread out a set of scores is. If a set of scores has a mean of 10 and a standard deviation of 1, most of the scores in that set will be very close to one another and to the mean of 10. About two-thirds of the scores will be between 9 and 11. A group whose scores are close together is called *homogeneous*. A group in which the scores are much more spread out, where the standard deviation is much larger, is called *heterogeneous*. For example, a set of scores with a mean of 10 and a standard deviation of 5 is more spread out than the first example (it is therefore heterogeneous); about two-thirds of this group's scores would be between 5 and 15.

Other descriptive statistics include *stanines*, *percentiles*, and *standard* scores. Stanines break the distribution of scores into nine sections (1=lowest, 9=highest) and indicate in which of the nine sections an individual score falls. Percentiles (from 0 to 99) describe the percentage of scores that are lower than an individual score. Standard scores describe how far an individual score is from the mean score: if 0 represents the mean, then a standard score of 1.5 means that the individual score is one and one-half standard deviations above the mean, whereas a score of minus 2 means a score is two standard deviations below the mean.

Inferential Statistics

Whereas descriptive statistics describe what the information is, *inferential statistics* tell what can be implied from that information. Inferential statistics tell us the odds in which the differences between groups can be attributed to chance or are real and could be replicated (that is, if the study were done again, the results would be similar).

Most researchers feel that they have to be at least 95 percent certain that differences between groups are real before they are willing to say so. This is known as the *level of probability* or *significance* and is generally shown as $p < .05$, meaning that the differences between groups were large enough that the chances are better

than 95 out of 100 that the study could be replicated with similar results. The statement ($p < .05$) is considered an acceptable level of risk, and differences between groups at the $p < .05$ level are considered statistically significant. If the study reports ($p < .01$), then the chances are better than 99 out of 100; ($p < .001$) means the chances are better than 999 out of 1000; and so on.

Significant Differences

There are two types of significant differences: *statistical* and *practical*. Statistical significance means that the differences between or among groups are most likely real and would be found if the study were replicated. However, just because a difference is statistically significant does not necessarily mean that it has practical meaning. Statistical significance is related to a number of things, including the size of the differences between groups, the number of subjects, and the degree that the scores in each group are spread out. For example, a study of 10,000 students might find that students using Math Book A increased their math achievement 1 percent more than students using Math Book B did. This difference would be statistically significant, meaning that the differences were most likely real and not due to chance. A teacher choosing a math book, however, could and most likely would say that a 1 percent difference was not meaningful; with such a small difference, factors such as cost and ease of use would and should be the deciding factors in book selection. Thus the statistically significant difference would have no practical significance.

This has been only an introduction to research methods; there is much more to learn. For additional information, see F. Kerlinger's *Foundations of Behavioral Research* (New York: Macmillan, 1972) or any of the many educational research books available in local college or public libraries.

References

Altus, G. (1953). WISC patterns of a selective sample of bilingual school children. *Journal of Genetic Psychology*, 83, 241-48.

Anastasi, A., & Cordova, F. (1953). Some effects of bilingualism upon the intelligence test performance of Puerto Rican children in New York City. *Journal of Educational Psychology*, 45, 1-17.

Atkinson, J. (1958). *Motives in Fantasy, Action and Society*. Princeton, NJ: Van Nostrand.

Bean, R. (1906). Some racial peculiarities of the Negro brain. *American Journal of Anatomy*, 5, 353-432.

Becker, J. (1981). Differential treatment of females and males in mathematics classes. *Journal of Research in Mathematics Education*, 12, 40-53.

Becker, J., & Jacobs, J. (1983). Sex: Is it an issue in mathematics? *Educational Horizons*, Winter, 60-66.

Benbow, C., & Stanley, J. (1980). Sex differences in mathematical ability: Fact or artifact? *Science*, 210, 1262-64.

Benbow, C., & Stanley, J. (1983). Differential course-taking hypotheses revisited. *American Educational Research Journal*, 20, 469-73.

Best, R. (1983). *We All Have Scars*. Bloomington, IN: Indiana University Press.

Bettleheim, B., & Janowitz, M. (1964). *Social Change and Prejudice*. New York: Free Press.

Bevis, W.M. (1921). Psychological traits of the southern Negro with observations as to some of his psychoses. *American Journal of Psychiatry*, 1, 69-78.

Billingsley, A. (1968). *Black Families in White America*. Englewood Cliffs, NJ: Prentice-Hall.

Bliss, W. (1870). *Woman and Her Thirty Years' Pilgrimage*. Boston: B.B. Russell.

Bordelon, K. (1985). Sexism in reading materials. *The Reading Teacher*, 38, 792-97.

Brown, B. (1967). *The Assessment of Self-concept among Four-year-old Negro and White Children: A Comparative Study Using the Brown-IDS Self-concept Referents Test*. New York: Institute for Developmental Studies.

Bryan, T. (1976). Peer popularity of learning disabled children: A replication. *Journal of Learning Disabilities*, 9, 49-53.

Bullough, V., & Bullough, B. (1973). *The Subordinate Sex: A History of Attitudes toward Women*. Urbana, IL: The University of Illinois Press.

Campbell, P. (1980). Student interaction: The importance of race and sex. Paper presented to the American Educational Research Association, Boston, MA.

Campbell, P. (1981). *The Impact of Societal Biases on Research*. Washington, DC: National Institute of Education.

Campbell, P., & Brown, M. (1982). *Surveying the Status of Educational Researchers.* Washington, DC: National Institute of Education.

Campbell, P., & Klein, S. (1982). Equity issues in education. In H. Mitzel (ed.), *The Encyclopedia of Educational Research.* New York: Macmillan Company, 581-87.

Campbell, P., & Scott, E. (1980). Non-biased tests can change the score. *Interracial Books for Children Bulletin*, 11, 7-9.

Caplan, N., & Nelson, S. (1973). On being useful: The nature and consequences of psychological research of social problems. *American Psychologist*, 28, 119-211.

Caplan, P. (1975). Sex differences in anti-social behavior: Does research methodology produce or abolish them? *Human Development*, 6, 444-60.

Caplan, P.; MacPherson, G.; & Tobin, P. (1985). Do sex-related differences in spatial abilities exist? A multilevel critique with new data. *American Psychologist*, 40, 786-99.

Christiansen, T., & Livermore, G. (1970). A comparison of Anglo-American and Spanish American children on the WISC. *Journal of Social Psychology*, 81, 9-14.

Clarke, E. (1873, 1972). *Sex in Education; or, A Fair Chance for Girls.* Boston: James R. Osgood. (Reprint edition by Arno Press.)

Coffman, W. (1961). Sex differences in response to items in an achievement test. *Eighteenth Yearbook, National Council on Measurement in Education*, 18, 117-24.

Collis, B. (1985). Psychosocial implications of sex differences in attitudes toward computers: Results of a survey. *International Journal of Women's Studies*, 3, 207-13.

Committee on the Status of Women in Sociology (1980). Sexist bias in sociological research. *ASA Footnotes*, January, 6, 7, 9.

Cowan, M.; Watkins, B.; & Davis, W. (1975). Level of education, diagnosis and race-related differences in MMPI performance. *Journal of Clinical Psychology*, 31, 442-44.

Crawford, J. (1986). Education key to narrowed Black-white wage gap, study says. *Education Week*, March 5, 25.

Cross, A. (1984). *Sweet Death, Kind Death.* New York: Ballantine.

Donlon, T. (1971). Content factors in sex differences on test questions. Paper presented to the annual meeting of the American Educational Research Association, Boston, MA.

Dwyer, C. (1976). Test content and sex differences in reading. *Reading Teacher*, 29, 753-57.

Eagly, A., & Carli, L. (1981). Sex of researchers and sex-typed communications as determinants of sex differences in influenceability: A meta-analysis of social influence studies. *Psychological Bulletin*, 90, 1-20.

Ehrenreich, B., & English, D. (1979). *For Her Own Good.* Garden City, NY: Anchor Books.

Evarts, A. (1914). Dementia praecocs in the colored race. *The Psychoanalytic Review*, 1, 388-403.

Frauenheim, J. (1978). Academic achievement characteristics of adult males who were diagnosed as dyslexic in childhood. *Journal of Learning Disabilities*, 11, 21-28.

Fuchs, D.; Fuchs, A.; Dailey, A.; & Power, M. The effect of examiners' personal familiarity and professional experience on handicapped children's test performance. *Journal of Educational Research*, 78, 141-45.

Garrett, H. (1961). The equalitarian dogma. *Mankind Quarterly*, 1, 253-55.

Gerwitz, S., & Dodge, K. (1975). Adults' evaluation of a child as a function of the sex of the adult and the sex of the child. *Journal of Personality and Social Psychology*, 32, 822-28.

Gesell, A., & Ilg, F. (1943). *Infant and Child in the Culture of Today.* New York: Harper.

Gilligan, C. (1980). *In a Different Voice.* Cambridge, MA: Harvard University Press.

Gilligan, C. (1986). Changing the questions: Women's studies and psychological development. The inaugural speech of the Blanche, Edith and Irving Laurie New Jersey Chair in Women's Studies. Douglass College, Rutgers University, New Brunswick, NJ.

Goldberg, S., & Lewis, M. (1969). Play behavior in the year-old infant: Early sex differences. *Child Development*, 40, 21-31.

Gossett, T. (1963). *Race: The History of an Idea in America.* Dallas: Southern Methodist University Press.

Gould, S. (1981). *The Mismeasure of Man.* New York: W.W. Norton and Company.

Graves, Z., & Graves, J. (1978). The effect of context on mother-child interaction: A progress report. *The Quarterly Newsletter of the Institute for Comparative Human Development*, 41-46.

Greenberg, S. (1978). *Right from the Start.* Boston: Houghton Mifflin.

Greenwald, A. (1975). Consequences of prejudice against the null hypothesis. *Psychological Bulletin*, 82, 1-20.

Hall, G. (1905). *Adolescence Vol. II.* New York: D.D. Appleton.

Hall, J. (1978). Gender effects of decoding nonverbal cues. *Psychological Bulletin*, 85, 845-57.

Hearnshaw, L.S. (1979). *Cyril Burt, Psychologist.* London: Hodder and Stoughton.

Herman, L., & Serbin, L. (1977). Observer expectations of a child's performance as a function of sex. Paper presented to the Third Annual Midwest Conference on Research on Women in Education, Milwaukee, WI.

Herschberger, R. (1948). *Adam's Rib*. New York: Pellegrini and Cudhy.

Issac, S. (1975). *Handbook in Research and Evaluation*. San Diego: EdITS.

Jacobs, J., & Eccles, J. (1985). Gender differences in math ability: The impact of media reports on parents. *Educational Researcher*, 14, 20-24.

Jenson, A. (1969). How much can we boost IQ and achievement? *Harvard Educational Review*, 39, 1-139.

The Joint Committee on Standards for Educational Evaluation (1981). *Standards for Evaluation of Educational Programs, Projects, and Materials*. New York: McGraw-Hill.

Jones, L. (1986). The influence on mathematics test scores, by ethnicity and sex, of prior achievement and high school mathematics courses. *Journal of Research in Mathematics Education*, 17.

Jones, L.; Burton, N.; & Davenport, E. (1984). Monitoring the mathematics achievement of Black students. *Journal of Research in Mathematics Education*, 15, 154-64.

Kaufman, D., & Richardson, B. (1982). *Achievement and Women*. New York: Free Press.

Kerlinger, F. (1972). *Foundations of Behavioral Research*. New York: Macmillan.

Kohlberg, L., & Kramer, R. (1969). Continuities and discontinuities in childhood and adult moral development. *Human Development*, 12, 93-120.

Lewontin, R.; Rose, S.; & Kamin, L. (1984). *Not in Our Genes*. New York: Pantheon Books.

Lipson, D. (1986). Personal communication, March 19.

Maccoby, E., & Jacklin, C. (1974). *The Psychology of Sex Differences*. Stanford, CA: Stanford University Press.

Mathews, W. (1984). Editorial. *Journal of Research in Mathematics Education*, 5, 165-71.

McClelland, D. (1975). *Power: The Inner Experience*. New York: Irvington Publishers.

McDill, E.; Meyers, E.; & Rigsby, L. (1966). *Sources of Educational Climates in High Schools*. Baltimore: Johns Hopkins.

Nage, E. (1961) *The Structure of Science*. New York: Harcourt, Brace and World.

National Academy of Education (1984). *Improving Education: Perspectives on Educational Research*. Pittsburgh: National Academy of Education.

Nichols, P. (1978). Dynamic variation theory as a model for the study of language and sex. Paper presented at the Ninth World Congress of Sociology, Uppsala, Sweden.

Oakley, A. (1977). *The Sociology of Housework*. New York: Pantheon Books.

Pallas, A., & Alexander, K. (1983). Reply to Benbow and Stanley. *American Educational Research Journal*, 20, 475-77.

Parlee, M. (1975). Psychology. *SIGNS: A Journal of Women in Culture and Society*, 1, 119-38.

Petroni, F.; Hirsch, E.; & Petroni, C. (1970). *2 4 6 8 When We Gonna Integrate?* New York: Behavioral Publications.

Pettigrew, T. (1964). Negro-American personality: Why isn't more known? *Journal of Social Issues*, 20, 4-23.

Prescott, S., & Foster, K. (1974). Why researchers don't study women: The response of 67 researchers. Paper presented to the annual meeting of the American Psychological Association, New Orleans, LA.

Rosenthal, R., & Rubin, D. (1982). Further meta-analytic procedures for assessing cognitive gender differences. *Journal of Educational Psychology*, 74, 708-12.

Russell, B. (1959). *The Scientific Outlook*. New York: W.W. Norton.

Samuel, W. (1977). Observed IQ as a function of test atmosphere, tester expectation and race of tester: A replication for female subjects. *Journal of Experimental Psychology*, 69, 593-604.

Sattler, J. (1970). Racial experimenter effects in experimentation, testing, interviewing and psychotherapy. *Psychological Bulletin*, 72, 137-60.

Schofield, J., & Sanger, H. (1977). Peer interaction patterns in an integrated middle school. *Sociometry*, 40, 39-46.

Scott, P. (1982). Debunking Sapphire: Toward a non-racist and non-sexist social science. In G. Hull, P. Scott, and B. Smith (eds.), *But Some of Us Are Brave*. New York: Harper & Row, 85-92.

Scott-Jones, D., & Clark, M. (1986). The school experiences of Black girls: The interaction of gender, race, and socio-economic status. *Phi Delta Kappan*, 67, 520-26.

Senk, S., & Usiskin, Z. (1983). Geometry proof writing: A new view of sex differences in mathematics ability. *American Journal of Education*, February, 187-201.

Sexton, P. (1969). *The Feminized Male*. New York: Vintage Books.

Shakeshaft, C. (1979). Dissertation research on women in academic administration: A synthesis of findings and paradigm for future research. Ph.D. dissertation, Texas A&M University.

Slavin, R., & Madden, N. (1979). School practices that improve race relations. *American Educational Research Journal*, 16, 23-28.

Smith, J., & Welch, F. (1986). *Closing the gap: Forty years of economic progress for Blacks*. Washington, DC: Rand Corporation.

Smith, M.L. (1980). Publication bias and meta-analysis. *Evaluation in Education*, 4, 22-24.

Steinem, G. (1980). Feminist notes. *MS*, 9, 98-103.

Stetson, E. (1982). Studying slavery: Some literary and pedagogical considerations on the Black female slave. In G. Hull, P. Scott, and B. Smith (eds.), *But Some of Us Are Brave*. New York: Harper & Row, 61-84.

Stivers, P., & Leckie, D. (1976). *NIE/AERA Planning Conference to Increase the Participation of Women and Minorities in Educational Research and Development*. Washington, DC: National Institute of Education.

Terman, L. (1916). *The Measurement of Intelligence*. Boston: Houghton Mifflin.

Thomas, A., & Sillen, S. (1972). *Racism and Psychiatry*. New York: Brunner/Mazel.

Tolbin, M. (1972). The Darwinian influence on psychological definitions of femininity in England. Doctoral dissertation, Harvard University.

Tyler, R., & White, S. (1978). Testing, teaching and learning. *Chairman's Report of a Conference on Research in Testing*. Washington, DC: U.S. Government Printing Office.

Ward, K., & Grant, L. (1984). The feminist critique and a decade of published research in sociology journals. A paper presented at the annual meeting of the American Sociological Association, San Antonio, TX.

Weinberg, M. (1977). *Minority Students: A Research Approach*. Washington, DC: U.S. Government Printing Office.

Welch, W.; Andersson, R.; & Harris, L. (1982). Effects of schooling on mathematics achievement. *American Educational Research Journal*, 19, 145-53.

Witelson, S. (1978). Quoted in *Psychology Today*, November, 51.

CODE NO.: 0691

We all know that sex and race bias in research has been a fact of history. In *The Hidden Discriminator: Sex and Race Bias in Educational Research,* author Patricia B. Campbell shows that bias in educational research is still alive and well. Using examples both from the past and the present, Campbell examines the myriad ways that bias can affect language: its influence on the researchers, the groups selected for study, the questions that are asked, the way the study is done, and the conclusions that are drawn. The result is a startling picture of the influences that shape the research that we use and that is used on us.

In language that is both lively and down-to-earth, Campbell not only describes the problem, but provides easy-to-follow guidelines for evaluating research and identifies "next steps" for reducing the incidence and effects of bias in research. An educational tool that shows us we can't responsibly accept research without asking some basic questions about it, this unique book is a must for all those who use or are affected by educational research, including teachers, parents, administrators, counselors, and students.

"*The Hidden Discriminator* provides an excellent discussion of race and sex bias in research methods currently not in textbooks on research."

— *Charol Shakeshaft, Hofstra University*